

## 本周周报（2013.10.14-2013.10.20）

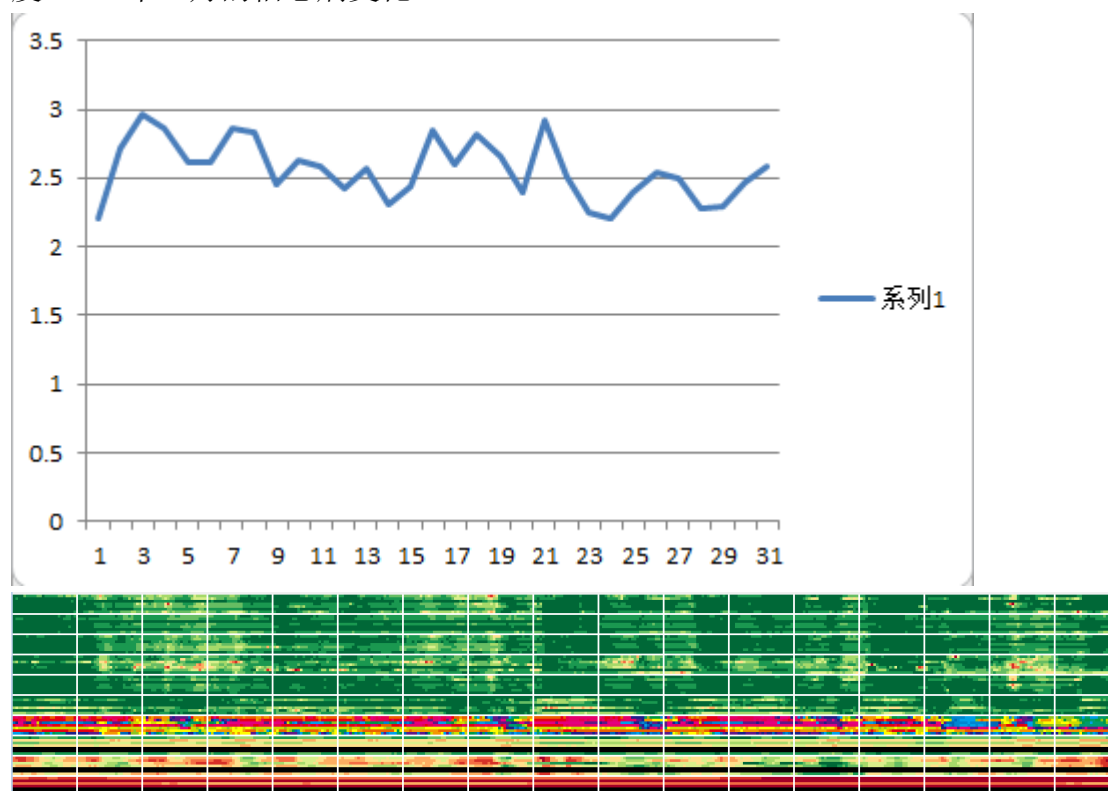
郭方舟

### 本周工作

#### 1. 空气污染数据可视化

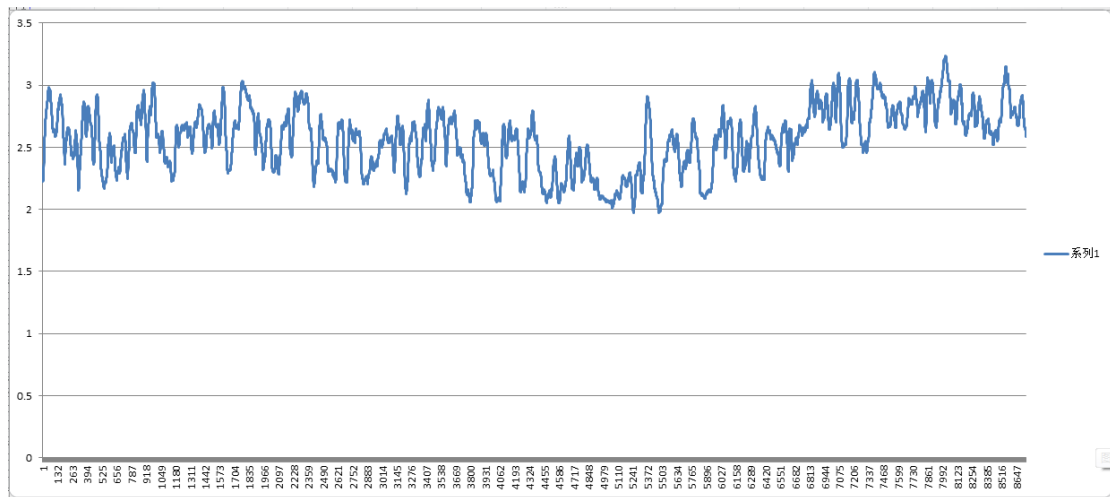
本周使用 R 语言中的 `cca` 包对属性间的相关性进行了分析，但是没有看到很好的结果，在跟海东师兄交流之后，感觉 `cca` 不是很适合这个数据，如果要计算相关性的话可以直接使用协方差矩阵。由于现在的进度还没有进行到相关性计算，所以这个问题暂时搁置。

上周提到的三个任务中，本周在进行第一个任务，即使用信息熵来标识事件。计算了以 1 天，3 天，1 周，2 周以及一个月为长度的信息熵，下图是以 1 天为长度 2009 年 1 月的信息熵变化：



在计算中没有包含风向这个属性，可以看到，信息熵的变化和像素图中变化是一致的，要识别事件最简单的方法就是对信息熵设定一个阈值，高于阈值即可标明为一个事件。

跟斐然师兄讨论后，斐然师兄想要将熵的计算结果做成一种堆叠效果，这需要在计算熵的过程中加窗，目前实现了以一天为窗长，一小时为窗移的信息熵计算，结果如下：



上图显示了一年中的熵变化，其变化趋势与像素图的结果吻合。  
在将熵变化做成堆叠效果后，可以添加交互，通过交互的方式决定哪些时间点使用哪一个层级的窗，将像素图部分合并，这样可以达成突出了事件的目标。

查阅了互信息的含义，互信息是基于信息熵的，是一种计算相关性的算法。在找出时间之后可以尝试这个算法进行相关性计算。

## 2. 读《Visualization Design and Analysis: Abstractions, Principles, and Methods》和《Grammar of Graphics》

第一本书读完了前两章，第二本书读到第二章的一半，读书笔记正在整理中。

## 3. 数据堂的数据已经收到并上传到服务器上

### 下周工作

#### 1. 空气污染数据的可视化

下周报告的论文《Visual Compression of Workflow Visualizations with Automated Detection of Macro Motifs》与这个项目之间可能有一定的联系，特别是自动识别 macro motifs 这一点，读完论文后想一想这篇文章使用的方法在空气污染数据中的事件识别联系。

完成熵堆叠三角的工作，看看画出来的效果是否真的能辅助事件识别。

#### 2. 读

#### 3. 完成秋学期课程作业